RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: _____ 10/031, 496 E_____

Source: _____ IFW16_____

Date Processed by STIC: _____ 03/22/2007_____

ENTERED

```
         RAW SEQUENCE LISTING               DATE: 03/22/2007
         PATENT APPLICATION:  US/10/031,496E      TIME: 12:58:50

         Input Set : N:\efs\03_22_07\10031496e_efs\99-45_SEQUENCE_ST25.txt
         Output Set: N:\CRF4\03222007\J031496E.raw

    3 <110> APPLICANT: National Renewable Energy Laboratory (NREL)
    5 <120> TITLE OF INVENTION: Cellobiohydrolase I Gene and Improved Variants
    7 <130> FILE REFERENCE: NREL 99-45
    9 <140> CURRENT APPLICATION NUMBER: 10/031,496E
   10 <141> CURRENT FILING DATE: 2002-01-14
   12 <160> NUMBER OF SEQ ID NOS: 97
   14 <170> SOFTWARE: PatentIn version 3.4
   16 <210> SEQ ID NO: 1
   17 <211> LENGTH: 45
   18 <212> TYPE: DNA
   19 <213> ORGANISM: Artificial
   21 <220> FEATURE:
   22 <223> OTHER INFORMATION: Nucleotide encoding linker
   24 <400> SEQUENCE: 1
   25 cctcccggcg gaaacccgcc tggcaccacc accacccgcc gccca                45
   28 <210> SEQ ID NO: 2
   29 <211> LENGTH: 15
   30 <212> TYPE: PRT
   31 <213> ORGANISM: Trichoderma reesei
   33 <400> SEQUENCE: 2
   35 Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr Thr Arg Arg Pro
   36 1               5                   10                  15
   39 <210> SEQ ID NO: 3
   40 <211> LENGTH: 24
   41 <212> TYPE: DNA
   42 <213> ORGANISM: Artificial sequence
   44 <220> FEATURE:
   45 <223> OTHER INFORMATION: Nucleotide encoding linker
   47 <400> SEQUENCE: 3
   48 ggcggaaacc cgcctggcac cacc                                      24
   51 <210> SEQ ID NO: 4
   52 <211> LENGTH: 1551
   53 <212> TYPE: DNA
   54 <213> ORGANISM: Trichoderma reesei
   56 <400> SEQUENCE: 4
   57 atgtatcgga agttggccgt catctcggcc ttcttggcca cagctcgtgc tcagtcggcc    60
   59 tgcactctcc aatcggagac tcacccgcct ctgacatggc agaaatgctc gtctggtggc   120
   61 acgtgcactc aacagacagg ctccgtggtc atcgacgcca actggcgctg gactcacgct   180
   63 acgaacagca gcacgaactg ctacgatggc aacacttgga gctcgaccct atgtcctgac   240
   65 aacgagacct gcgcgaagaa ctgctgtctg gacggtgccg cctacgcgtc acgtacgga   300
   67 gttaccacga gcggtaacag cctctccatt ggctttgtca cccagtctgc gcagaagaac   360
   69 gttggcgctc gcctttacct tatggcgagc gacacgacct accaggaatt caccctgctt   420
   71 ggcaacgagt ctctttcga tgttgatgtt tcgcagctgc cgtgcggctt gaacggagct   480
```

RAW SEQUENCE LISTING   DATE: 03/22/2007
PATENT APPLICATION: US/10/031,496E   TIME: 12:58:50

Input Set : N:\efs\03_22_07\10031496e_efs\99-45_SEQUENCE_ST25.txt
Output Set: N:\CRF4\03222007\J031496E.raw

```
 73 ctctacttcg tgtccatgga cgcggatggt ggcgtgagca agtatcccac caacaccgct    540
 75 ggcgccaagt acggcacggg gtactgtgac agccagtgtc cccgcgatct gaagttcatc    600
 77 aatggccagg ccaacgttga gggctgggag ccgtcatcca acaacgcgaa cacgggcatt    660
 79 ggaggacacg gaagctgctg ctctgagatg gatatctggg aggccaactc catctccgag    720
 81 gctcttaccc cccacccttg cacgactgtc ggccaggaga tctgcgaggg tgatgggtgc    780
 83 ggcggaactt actccgataa cagatatggc ggcacttgcg atcccgatgg ctgcgactgg    840
 85 aacccatacc gcctgggcaa caccagcttc tacggccctg gctcaagctt taccctcgat    900
 87 accaccaaga aattgaccgt tgtcacccag ttcgagacgt cgggtgccat caaccgatac    960
 89 tatgtccaga atggcgtcac tttccagcag cccaacgccg agcttggtag ttactctggc   1020
 91 aacgagctca acgatgatta ctgcacagct gaggaggcag aattcggcgg atcctctttc   1080
 93 tcagacaagg gcggcctgac tcagttcaag aaggctacct ctggcggcat ggttctggtc   1140
 95 atgagtctgt gggatgatta ctacgccaac atgctgtggc tggactccac ctacccgaca   1200
 97 aacgagacct cctccacacc cggtgccgtg cgcggaagct gctccaccag ctccggtgtc   1260
 99 cctgctcagg tcgaatctca gtctcccaac gccaaggtca ccttctccaa catcaagttc   1320
101 ggacccattg gcagcaccgg caaccctagc ggcggcaacc ctcccggcgg aaacccgcct   1380
103 ggcaccacca ccacccgccg cccagccact accactggaa gctctcccgg acctacccag   1440
105 tctcactacg gccagtgcgg cggtattggc tacagcggcc ccacggtctg cgccagcggc   1500
107 acaacttgcc aggtcctgaa cccttactac tctcagtgcc tgtaaagctc c            1551
110 <210> SEQ ID NO: 5
111 <211> LENGTH: 514
112 <212> TYPE: PRT
113 <213> ORGANISM: Trichoderma reesei
115 <400> SEQUENCE: 5
117 Met Tyr Arg Lys Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Arg
118 1               5                   10                  15
121 Ala Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr
122             20                  25                  30
125 Trp Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser
126         35                  40                  45
129 Val Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser
130     50                  55                  60
133 Thr Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp
134 65                  70                  75                  80
137 Asn Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala
138                 85                  90                  95
141 Ser Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe
142             100                 105                 110
145 Val Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met
146         115                 120                 125
149 Ala Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe
150     130                 135                 140
153 Ser Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala
154 145                 150                 155                 160
157 Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro
158                 165                 170                 175
161 Thr Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln
162             180                 185                 190
165 Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly
166         195                 200                 205
```

```
169 Trp Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly
170     210             215             220
173 Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu
174 225             230             235             240
177 Ala Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu
178             245             250             255
181 Gly Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr
182         260             265             270
185 Cys Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr
186         275             280             285
189 Ser Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys
190     290             295             300
193 Leu Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr
194 305             310             315             320
197 Tyr Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly
198             325             330             335
201 Ser Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu
202         340             345             350
205 Ala Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln
206     355             360             365
209 Phe Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp
210     370             375             380
213 Asp Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr
214 385             390             395             400
217 Asn Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr
218             405             410             415
221 Ser Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys
222 ·         420             425             430
225 Val Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn
226         435             440             445
229 Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr
230     450             455             460
233 Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln
234 465             470             475             480
237 Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val
238             485             490             495
241 Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln
242         500             505             510
245 Cys Leu
249 <210> SEQ ID NO: 6
250 <211> LENGTH: 514
251 <212> TYPE: PRT
252 <213> ORGANISM: Trichoderma reesei CBH1-N45A
254 <400> SEQUENCE: 6
256 Met Tyr Arg Lys Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Arg
257 1           5               10              15
260 Ala Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr
261             20              25              30
264 Trp Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser
```

```
265          35                    40                   45
268 Val Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Ala Ser Ser
269      50                    55                   60
272 Thr Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp
273 65                  `70              75                  80
276 Asn Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala
277                 85              90                  95
280 Ser Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe
281             100             105             110
284 Val Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met
285         115             120             125
288 Ala Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe
289         130             135             140
292 Ser Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala
293 145                 150             155                 160
296 Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro
297                 165             170                 175
300 Thr Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln
301             .180            185             190
304 Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly
305         195             200             205
308 Trp Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly
309     210                 215             220
312 Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu
313 225             230             235                 240
316 Ala Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu
317             245             .  250            255
320 Gly Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr
321         260             265             270
324 Cys Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr
325         275             280             285
328 Ser Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys
329     290                 295             300
332 Leu Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr
333 305             310             315                 320
336 Tyr Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly
337             325             330             335
340 Ser Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu
341             340             345             350
344 Ala Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln
345         355             360             365
348 Phe Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp
349     370                 375             380
352 Asp Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr
353 385                 390             395                 400
356 Asn Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr
357             405             410 .           415
360 Ser Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys
361             420             425             430
```

```
364 Val Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn
365         435             440                 445
368 Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr
369     450             455                 460
372 Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln
373 465             470                 475                 480
376 Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val
377         485             490                 495
380 Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln
381         500             505                 510
384 Cys Leu
388 <210> SEQ ID NO: 7
389 <211> LENGTH: 514
390 <212> TYPE: PRT
391 <213> ORGANISM: Trichoderma reesei CBH1-N270A
393 <400> SEQUENCE: 7
395 Met Tyr Arg Lys Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Arg
396 1               5               10                  15
399 Ala Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr
400         20              25                  30
403 Trp Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser
404     35              40                  45
407 Val Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser
408     50              55                  60
411 Thr Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp
412 65              70              75                  80
415 Asn Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala
416             85              90                  95
419 Ser Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe
420         100             105                 110
423 Val Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met
424         115             120                 125
427 Ala Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe
428     130             135                 140
431 Ser Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala
432 145             150             155                 160
435 Leu Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro
436         165             170                 175
439 Thr Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln
440         180             185                 190
443 Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly
444         195             200                 205
447 Trp Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly
448     210             215                 220
451 Ser Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu
452 225             230             235                 240
455 Ala Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu
456         245             250                 255
459 Gly Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 03/22/2007
PATENT APPLICATION: US/10/031,496E   TIME: 12:58:51

Input Set : N:\efs\03_22_07\10031496e_efs\99-45_SEQUENCE_ST25.txt
Output Set: N:\CRF4\03222007\J031496E.raw

Invalid <213> Response:
Use of "Artificial" only as "<213> Organism" response is incomplete,
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:1,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32
Seq#:33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54,55,56
Seq#:57,58,59,60,61,62,63,64,65,66,67,68,69,70,71,72,73,74,75,76,77,78,79,80
Seq#:81,82,83,84,85,86,87,88,89,90,91,92,93,94,95,96,97

VERIFICATION SUMMARY   DATE: 03/22/2007
PATENT APPLICATION: US/10/031,496E   TIME: 12:58:51

Input Set : **N:\efs\03_22_07\10031496e_efs\99-45_SEQUENCE_ST25.txt**
Output Set: **N:\CRF4\03222007\J031496E.raw**